

<https://helda.helsinki.fi>

Assessing historical literacy among 12-year-old Finns

Rantala, Jukka

2018-08-13

Rantala , J & Khawaja , A 2018 , ' Assessing historical literacy among 12-year-old Finns ' ,
The Curriculum journal , vol. 29 , no. 3 , pp. 354-369 . <https://doi.org/10.1080/09585176.2018.1460273>

<http://hdl.handle.net/10138/305943>

<https://doi.org/10.1080/09585176.2018.1460273>

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Jukka Rantala and Amna Khawaja

Assessing historical literacy among 12-year-old Finns

The new national core curriculum has been put into operation in Finland. Teaching and learning is intended to focus on historical literacy. In this paper, we study how it can be assessed with a pen-and-paper test, and what that reveals in regards to the mastering of historical literacy among elementary pupils. To study this, we designed a test where pupils analysed sources, answered open-ended questions, and filled out weighted multiple-choice questions. Furthermore, we implemented the test with the think-aloud method. The results from this indicate the way 12-year-olds read historical documents and formulate their own interpretations. The results demonstrate that most of the pupils succeeded in acquiring historical information from different sources, as well as understanding the interpretative nature of historical knowledge. The test also revealed that pupils can identify some intentions of the sources' producers, as well as reasoning through their own interpretations using a single source. Some pupils, however, could not process the cognitive noise inherent to the test (i.e. conflicting sources). Some of them also considered an authoritative source to be principally more reliable than other sources. Our study sheds light on which kind of task can be used at the elementary level to assess historical literacy.

Keywords: History education; historical literacy; assessing historical literacy; elementary pupils; disciplinary knowledge

Introduction: Why do the Finns need material for criteria-based assessment in basic education?

It is obligatory in Finland for each child 7 to 16 years of age to complete their 9 years of comprehensive basic schooling. This is comprised partly of history, which is taught to students during their primary schooling and lower secondary schooling (Grades 5 and 6 for 10- to 12-year-olds, and Grades 7 and 8 for 13 to 15-year-olds). Concerning Finnish teachers, the upper administration of centralised education does not play a large role in their everyday jobs, meaning that the teachers have a great deal of autonomy. In Finland, there are no national examinations at the end of basic education, school inspectorate systems, nor formal teacher evaluations. Due to these aspects, it is left to the discretion of the teacher as an individual whether or not they develop their teaching based on the national core curriculum.

The core curriculum conducted by the Finnish National Board of Education consists of two parts. The general portion presents the mission and general goals of education. As for the disciplinary portion, it reveals the objectives and content areas.

Concerning the latter, teachers can find the assessment criteria that are common to all schools and teachers.

The fundamental problem within the Finnish system, in our opinion, is that the assessment criteria for history—as well as for other subjects—at the end of Grades 6 and 8/9 are determined only at the level of ‘good’ competence (numerical grade 8), but not at the other levels of competence (‘weak’, ‘passable’, ‘satisfactory’, ‘very good’, and ‘excellent’, represented by numerical grades from 4 through 10). The core curriculum guides teachers to use the national assessment criteria when determining the achievement levels of their students. We refer to the assessment criteria for history at the end of Grade 6 at the core curriculum in order to show that this is a concrete problem. An equivalent description is stated at the end of Grade 8. The definition of mastery of good knowledge and skills for both is defined loosely, and therefore is difficult for teachers to understand the border between, for example, an evaluation of ‘good’ for grade 8 and ‘very good’ for grade 9, since they don’t have any tools for calibrating their evaluation on a wider context. For example, there are no descriptions of a typical answer of students for each numerical grade on each objective stated in the core curriculum. Therefore, it is difficult for teachers to appraise the abilities of their students in comparison with the national level.

As stated earlier, students are not required to submit to a standardised, high-stakes test as part of completing their comprehensive schooling. The students are given grades by the teachers of their own school, and are able to apply for admission to either general upper secondary school or vocational upper secondary school with these grades.

In order to ensure that the students are being graded fairly, the Finnish National Board of Education carries out assessments of learning outcomes on a large scale and distributes the results of these sample-based assessments to schools and teachers (the result of which is not intended for ranking schools). This national analysis is carried out regularly using samples as a means of assessment. This attempts to ensure consistency regarding teacher-based grading (National Board of Education, 1999). Although schools are able to gauge their success based on the national average, the results of these assessments are not publicly accessible and are used only for schools’ and teachers’ self-evaluation. Therefore, discretion is left to the individual teacher as to what they do with this information.

These kinds of sample-based assessments carried out by the National Board of Education have usually been directed towards reading, mathematics, and science. Thus far, the first and only assessment aimed at students’ knowledge of history and social studies was carried out in the spring of 2011, based on the study of 4726 students (2352 boys and 2370 girls) from a student population comprising roughly 8 percent of their age group (the total being 60,000). The design of this test was aimed at gauging students’ knowledge in comparison with the core curriculum’s final assessment criteria.

Written instructions, explaining the requirements for specific scores, were laid out in detail and followed by the teachers implementing the test. Examples were provided in order to further aid teachers in their evaluation. In addition, five external censors were chosen by the Finnish National Board of Education to assess 694 (15%) of the students’ answers. Since the censors and teachers arrived at similar result, it can be

concluded that the evaluation instructions were straightforward (Ouakrim-Soivio & Kuusela, 2011).

This test in history and social studies—much like the tests in other school subjects—revealed that teachers do not have enough guidance for the evaluation of the mastery of students’ disciplinary knowledge (Hildén, Ouakrim-Soivio & Rautopuro, 2016; Kuusela, 2006; Ouakrim-Soivio, 2013). When comparing the results of the students who participated in the 2011 history and social studies test with that of the grades given to them by their teachers outside this test, we find that the evaluation of the students is not uniform. In some schools, teachers assess their students more loosely and in other schools more strictly, as can be seen in Figure 1.

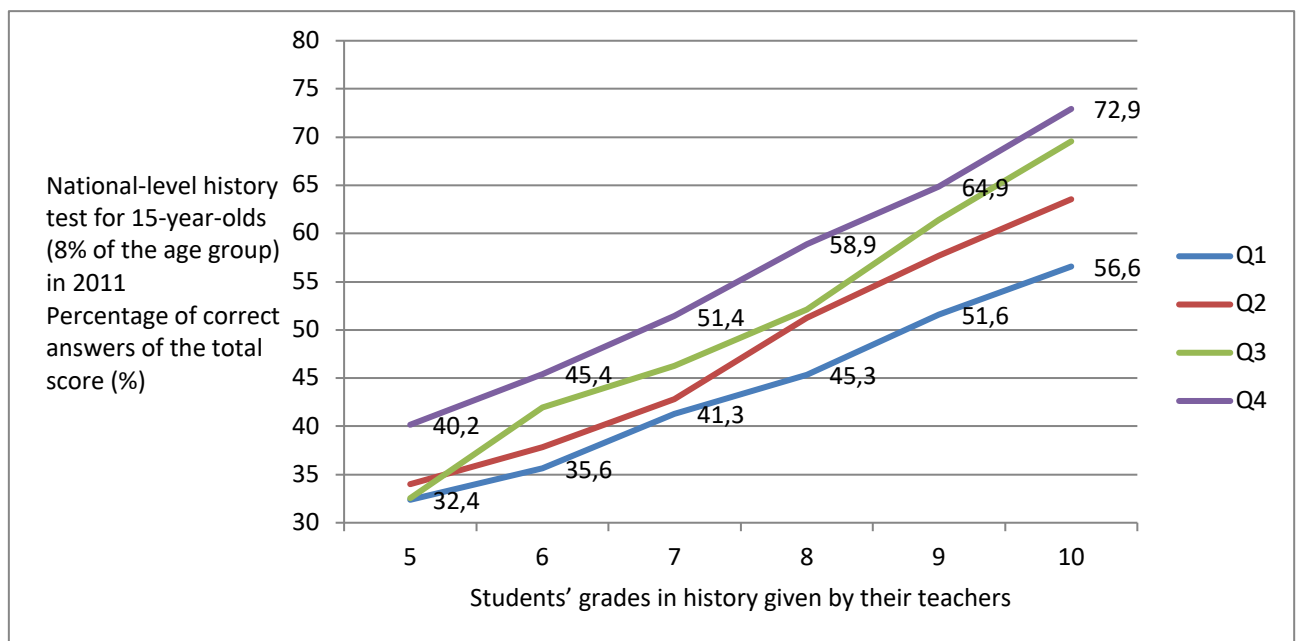


Figure 1. The assessment differences in schools between the grades given to students and the competencies revealed by the assessment of learning outcomes conducted by the Finnish National Board of Education in 2011. Source: Ouakrim-Soivio (2013, p. 176).

The figure shows how the students at the same level of competence received different grades. Ouakrim-Soivio and Kuusela (2012) discovered that teachers adjust their grades according to the general competence level of their class. They classified schools by the given grades into four categories, from the loosest (Q1) to the strictest (Q4). They showed that with the same competence level, students could even have two grade variations; for example, students scoring 40 percent of the maximum in the national-level history test conducted by the National Board of Education had school grades ranging from 5 to 7. Ouakrim-Soivio (2013) has stated that the core curriculum should provide teachers with a more defined and coherent grading assessment criteria and material for criteria-based evaluation. Ouakrim-Soivio and Kuusela’s (2012) study revealed that in 2011, one third of history teachers considered the assessment criteria to

be insufficient. Similar results have been found in recent studies (for example, Ouakrim-Soivio & Rantala, 2017).

Why does the National Board not offer a defined assessment criteria to teachers? Finnish basic education emphasises school-based assessment, in which teachers evaluate the students' learning processes and development throughout their school career. With the help of the results of their assessment, teachers can change their instruction and teaching methods. (Jakku-Sihvonen, 2001, p. 18; Jakku-Sihvonen & Heinonen, 2001, pp. 80–85; see also Brookhart, 2007, p. 43) Perhaps the National Board does not want to build an assessment system that would confine this, but rather might desire to keep assessment as a tool **for** learning instead **of** learning. The Board might also want to prevent a neo-liberalistic measuring culture in Finland, which before long could cause a segregation of those schools deemed high-performing versus those deemed low-performing. Presumably, the National Board of Education wants to prevent this in basic schooling, particularly because basic schooling is known to be of uniform quality of its results and the differences between schools are still very small (Kupari et al., 2013:42; Vettenranta et al., 2016, pp. 56–59; see also e.g. OECD, 2016, pp. 265–266).

Nevertheless, it is obvious that teachers need more guidance in regards to their assessment procedures, and there should be a more exact guidance for teachers assessing the mastering of history. This issue has motivated us to produce tests for teachers that help them to fairly assess their students' performance, as well as raising the equity of their assessment.

Historical literacy as a learning objective

History education today emphasises disciplinary literacy, meaning that students should have the abilities to process various texts like historians (Stahl & Shanahan, 2004; Wineburg, 2001). They should be able to scrutinise historical texts critically, evaluate their meanings, and use them to support their own interpretations of the past. They should also be able to contextualise texts in addition to having some sort of understanding about the author's position in society, intentions, and supposed audience (see Bain, 2006; Nokes, 2013, p. 20; Perfetti et al., 1994, pp. 260–262; Reisman, 2012; Wineburg & Reisman, 2015).

There are scholars who question young students' intellectual capacity investigating the past or abilities to think abstractly (see VanSledright, 2002, pp. 7–9). According to them, teaching historical literacy would be time misspent. On the other hand, the research of VanSledright (2002) and Lee and Ashby (2000), for instance, indicates that 12-year-olds are indeed capable of historical thinking. Additionally, Downey and Long (2016, p. 7) have stated that “historical literacy is a reasonable goal for elementary students, when outcomes are adjusted to the appropriate age level.”

There are also encouraging results from Sweden that show many elementary pupils being able to demonstrate at least some aspect of disciplinary thinking in history. Samuelsson and Wendell (2016) studied 12-year-old pupils' abilities in analysing historical sources in quite a similar context as we carried out our study. The history

curriculum for Grades 4–6 in Sweden expects that the pupils show disciplinary thinking, much in the same way as Finland.

In Finland, historical literacy is a fresh concept in the new national core curriculum. According to a general description of the task of history teaching, one objective of the instruction is to support the development of historical literacy. Historical literacy is often used synonymously with historical thinking: a historically-literate person has a grasp on both substantive and procedural knowledge, that is, he or she understands both historical context and knows how the historical knowledge regarding it is produced by someone. In addition, he or she also understands the difference in narratives surrounding the same historical event or phenomenon.

Even though historical literacy is a new concept in the core curriculum, Finnish history teaching already started to emphasise historical thinking at the beginning of the twenty-first century, and historical thinking was included in the National Core Curriculum and in the final assessment criteria in 2004. However, the assessment of learning outcomes in 2011 exposed that students' historical knowledge corresponds with the earlier content-based curriculum rather than the disciplinary-based curriculum, revealing that not all of the teachers were teaching in accordance with the present demands (Rantala, 2012). Similar observations concerning the lack of disciplinary aspects in teaching have been made in Sweden despite their disciplinary history curriculum (Thorp & Törnqvist, 2017, p. 215).

Both Finland and Sweden have changed their history curricula relatively recently. Pioneering countries in teaching and assessing historical literacy can be found in the Anglo-American world. For example, the Amherst Project in the USA in the 1950s, as the Schools Council History Project in the UK that followed two decades later were the first attempts to teach students to think historically. Recently, teaching and learning historical literacy has been studied in the UK (Lee & Ashby, 2000), Canada (Seixas & Colyer, 2011; Seixas & Morton, 2013), and the USA (Reisman, 2012). Nevertheless, there is scant attention paid towards elementary pupils' historical thinking. There are also different views among scholars regarding the age at which children are perceived capable of carrying out challenging interpretive tasks (see Rantala & van den Berg, 2015, p. 72).

Critical thinking, problem-solving, and literacy in different streams of information are all factors that correspond to skills needed in the twenty-first century, as well as that of teaching historical literacy (see, for example, Husbands, Kitson & Pendry, 2003; Nokes, 2013; Seixas & Ercikan, 2015). Historical literacy and historical thinking skills—such as the skills needed to compare documents and justify arguments—have been shown quite unanimously to have a close relationship with each other (Nokes, 2013; Perfetti et al., 1994, pp. 260–261; VanSledright & Frankes, 2000; Wineburg, 1991). Therefore, it is already vital at the elementary level to begin learning historical literacy.

In the Finnish national core curriculum for basic education, historical literacy is narrowly defined as “the ability to read and analyse sources produced by actors of the past and to competently interpret their meaning and significance” (National Board of Education, 2014:295–296). When comparing this definition with that of the assessment

criteria defined in Table 1, historical literacy can also include the following objectives: “to guide the pupil to recognise different sources of history,” “to guide the pupil to notice that historical information can be interpreted in different ways,” and “to instruct the pupil to explain how interpretations may change as a consequence of the new sources or new ways of examining them.”

Table 1. Assessment criteria for history for a verbal assessment describing good knowledge and skills at a Grade 6 level, assessed as an 8 (‘good’). Source: National Board of Education (2014, p. 298).

| Objective of instruction | Knowledge and skills for the verbal assessment ‘good’ (numerical grade 8) |
|--|---|
| <i>Significance, values, and attitudes</i> | |
| O1 to guide the pupil to become interested in history as a field of knowledge and a subject that builds his or her identity | The development of the pupil's motivation is not used as a basis for grade formulation. The pupils are guided in reflecting on their experiences as a part of self-assessment. |
| <i>Acquiring information about the past</i> | |
| O2 to guide the pupil to recognise different sources of history | The pupil is able to search for historical information in different sources with guidance. |
| O3 to guide the pupil to notice that historical information can be interpreted in different ways | The pupil is able to tell facts from interpretations. |
| <i>Understanding historical phenomena</i> | |
| O4 to help the pupil to understand different ways of dividing history into eras and using the related historical concepts | The pupil recognises the main ways of structuring time in history and is able to give examples of typical features of societies in different times and different eras. |
| O5 to guide the pupil to understand the motives of human activity | The pupil is able to put himself or herself in the position of a person of the past and to describe the motivations of his or her actions. |
| O6 to help the pupil to perceive different reasons for historical events and phenomena and their consequences | The pupil recognises and is able to give examples of causal relationships of historical phenomena. |
| O7 to help the pupil to identify changes in the history of his or her family or community and to understand how the same changes may have meant different things to different people | The pupil is able to describe changes and explain why change does not equal progress. Using some examples, the pupil is able to describe how the same change has had a different meaning for different people and groups. |
| O8 to teach the pupil to perceive the continuity in history | The pupil is able to give examples of the continuity of phenomena from one era to another. |
| <i>Applying historical knowledge</i> | |
| O9 to guide the pupil to propose reasons for changes | The pupil is able to describe the main features of the causal relationships of some historical phenomena. |
| O10 to instruct the pupil to explain how interpretations may change as a consequence of new sources or new ways of examining them | Using some examples, the pupil is able to explain why the same event or phenomenon may be interpreted in different ways. |
| O11 to guide the pupil to explain human activity | The pupil is able to describe the studied event or a phenomenon from the point of view of different actors. |

In the following, we portray what our task reveals about pupils’ mastery of those objectives.

Material and data gathering

Participants

The task we used in gathering data was put into practice in a mid-size comprehensive school in Helsinki in November of 2015. Eighteen pupils accomplished the task as a pen-and-paper test, and two pupils completed it using the think-aloud method. The think-aloud method was used mainly to scrutinise the validity of the pen-and-paper test.

According to the objectives of history teaching in the national core curriculum, elementary pupils should learn, compare, and form their own justified opinions based on using a variety of sources, as well as understanding that historical information can be interpreted in different ways. The teacher of our test class, who had a career as a teacher for 35 years, told us that she follows the directions of the core curriculum. She described to us how she had taught both substantive and procedural knowledge to her pupils, therefore it would not be the first time that the pupils were asked to interpret historical documents.

The pupils formed a typical 6th Grade class in Finland, totalling 23 pupils (though three were sick the day when our task was realised). This experienced teacher estimated her group as typical in its level of performance when compared to that of her previous classes and the other 6th Grade classes in her school.

Pen-and-paper task

We designed a pen-and-paper task four pages in length for assessing the mastery of historical literacy. The task was a typical source exercise, consisting of three written documents and two pictures. The topic of the exercise was relationship between Nicholas Copernicus and the Catholic Church.

We chose to work with this particular historical context because of its close connection to the National Core Curriculum, which states five content areas that should be studied during Grades 5 and 6. One of the content areas is entitled “The revolution of the modern times: The pupils are familiarised with changes that took place in science, arts, and people's beliefs.” Although it had been suggested already in Ancient Greek that the Earth was not the centre of the Universe, it was Copernicus who set things in motion for the acceptance of the heliocentric model.

Assessment tools for historical literacy have mostly been designed for secondary and upper secondary school level (cf. Reisman, 2012; Seixas, Gibson & Ercikan, 2015; VanSledright, 2014). As we were designing an assessment task for primary school pupils, we noticed that there was no equivalent material as a point of reference. We decided to build the assessment task around sources, which is a common approach when practicing historical thinking and historical literacy. The sources were modified in a way that would allow primary school pupils to properly understand them. This meant reducing the length considerably, as well as simplifying difficult phrases. According to Reisman (2012), it is acceptable and even necessary to modify sources for pupils even though the authenticity of the sources may be lost in the process.

Two pictures were used to find out pupils' mastery of the content knowledge needed to complete the task: namely, understanding the difference between geocentric and heliocentric models. The main task for the pupils was to explain why written sources differed in their framing of the relation between Copernicus and the Catholic Church. The sources are defined in Table 2. The purpose of our study was to find out how 12-year-old Finns master historical literacy. Our ultimate aim, however, is to build open-access tasks that help to better measure pupils' historical literacy through teachers' calibration of their assessments.

Table 2. Sources in our task

| Source | Description | Function in the task |
|--|--|--|
| Source 1. Geocentric model | An illustration of the Ptolemaic geocentric system by Bartolomeu Velho (1568) | Measuring the mastery of content knowledge needed to understand the task |
| Source 2. Heliocentric model | An illustration of the heliocentric system by Nicholas Copernicus (1543) | Measuring the mastery of content knowledge needed to understand the task |
| Source 3. Preface of Copernicus's book | Extract from the preface of Copernicus's book where he dedicates the book to the pope (1543); 83 words | Measuring historical literacy |
| Source 4. Cardinal's letter to Copernicus | Extract from a cardinal's letter to Copernicus where he encouraged Copernicus to publish his ideas (1536); 97 words | Measuring historical literacy |
| Source 5. Textbook passage | Extract from a textbook which tells that Copernicus attacked the Church and opposed the teachings of it (1982); 77 words | Measuring historical literacy |

In our task, there were two weighted multiple-choice and two open-ended questions, whose functions were to measure sourcing, contextualising and corroborating (see Wineburg, 1991). Our aim was to find out if pupils understood who wrote the texts and why (sourcing). We wanted to see whether or not they could connect texts to the timeframe in which they were written. We also wanted to see if pupils understood that Copernicus's book (source 3) and the Cardinal's letter (source 4) were situated in a time before the Catholic Church condemned the heliocentric model (contextualising). The most difficult task for young students, we expected, was to compare the sources against each other (corroborating). We built a contradiction between the sources to arouse the motivation of pupils to solve this 'cognitive noise'.

Data sources

The written documents were 257 words altogether, and the whole task was 707 words. Because participants were expected to complete a lot of reading, we wanted them to also produce a reasonable amount of writing. Therefore, we used weighted multiple-choice questions. The weighted multiple-choice questions were formed using the principles stated by VanSledright (2014): instead of one correct answer, there is one wrong answer and several possible answers which are weighted by their compelling possibility.

One of our multiple-choice questions asks why Copernicus dedicated his book to the Pope. There is one wrong answer ('he believed that the pope would not be interested in his book without it') and three possible acceptable options. One of them, however, is more compelling than the others (namely that 'Copernicus understood that his ideas might anger the Catholic Church, and that he wanted to stay in good standing with the Church'). The other possible acceptable options have some shortages. Though one of the options ('because he appreciated the Catholic Church so much') is essentially true, it does not explain why Copernicus dedicated his book to the Pope. The other option that is not very compelling ('he was afraid that the Pope would condemn his heliocentric model') reveals Copernicus' genuine worry. However, neither of these options are connected with a cause and consequence explanation detailing how Copernicus' act of dedicating the book to the Holy Father would soften the indignation within the Church.

Participants performing our task were also analysed using two open-ended questions. When creating the open-ended questions, we attempted to highlight the conflict in the sources we provided. The questions were asked in such way that forced participants to compare these sources: "When scrutinising the relationship between Copernicus and the Church, is the textbook as reliable a source as the sources that were written when Copernicus was alive?" and "The extract of the textbook says that Copernicus attacked the Church. Can you find something that supports or doesn't support this?" The open-ended questions were analysed with the help of criteria and model answers constructed by us. A parallel classifier was used ensuring the reliability of the assessment.

As one of the authors of this paper carried out the pen-and-paper test for eighteen pupils, the other author conducted a think-aloud session with two pupils. When using think-aloud method, participants attempted to verbalise their thoughts as accurately and simultaneously as possible (cf. Pressley & Afflerbach, 1995; Wineburg, 1991). The method has previously been used in research concerning problem-solving (Wineburg, 1991) and historical thinking (Breakstone, 2014). The teacher of the class selected the pupils for think-aloud sessions, and, according to her, the selection criterion was that they were not to be hesitant to speak with the researchers. Their think-aloud sessions were done one after another, and we prepared the participants for the think-aloud protocol before they accomplished the task. With the think-aloud method, we studied the consistency of our test, in other words how the questions in our test measured consistently historical literacy defined in the core curriculum.

Results

Weighted multiple-choice questions revealed how participants understood what they read, sorting strong from weak—as can be seen from Table 3.¹ However, these questions were not intended to reveal the historical thinking of the pupils. The other multiple-choice question (‘why Copernicus dedicated his book to the Pope’) revealed that many of the participants had difficulty defining the intentions of Copernicus. Almost half of them did not choose the wrong answer, but nevertheless chose the least possible answer (‘because he appreciated the Catholic Church so much’). Instead of disciplinary literacy, they utilised generic literacy: in the document, they recognised the respect Copernicus showed towards the pope. Their choice exposes that they did not recognise the hesitation of Copernicus and the context in which Copernicus’s book was published.

Table 3. An example of a weighted multiple-choice question: “How can you explain that the extract from a textbook gives a different kind of view than the extract from the cardinal’s letter on the reaction of the Church against the ideas of Copernicus?”

| | Score | Pupils no |
|--|-------|--------------|
| a) The textbook authors have used different sources that tell that the Church condemned the ideas of Copernicus, and they left out the sources that told the opposite story. | 4 | 14* |
| b) The cardinal was lying. | 0 | 1 |
| c) The textbook authors did not know that some churchmen supported the ideas of Copernicus. | 1 | 5 |
| d) The textbook authors did not find the cardinal’s letter an important source. | 2 | 1* |

* One pupil chose both the choices a and d.

Open-ended questions revealed how the participants interpreted the provided sources. Some pupils noticed a contradiction between the sources, yet could not solve it. Some pupils could not tolerate the cognitive noise inherent to the test (i.e. conflicting sources), and either did not recognise the conflicting information given by the sources or were unable to recognise the purpose of the source and consider the information to be a fact in each of the sources. In these cases, the next source seemed somehow to overwrite the previous one.

Some of the pupils considered an authoritative source to be principally more reliable than the other sources. Regarding the question “When scrutinising the relationship between Copernicus and the Church, is the textbook as reliable a source as the sources that were written when Copernicus was alive?” some pupils answered that you cannot put any nonsense in textbooks. Even though pupils understood the interpretative nature of other sources, they demonstrated that they thought a textbook

¹ The analysis is based on Rantala & Khawaja (2016).

could only report facts. Similar findings have been made elsewhere with older students (e.g. Wineburg, 1991).

These results demonstrated that some of the pupils succeeded in acquiring historical information from different sources, as well as understanding the interpretative nature of historical knowledge. The test also revealed that pupils could identify some of the intentions linked to the sources' producers, as well as reasoning through their own interpretations based on a single source. When they were asked to point out something from the other sources that would support or deny the content from the textbook that claimed that Copernicus attacked the Catholic Church, some pupils could give a solid answer:

In source 3, Copernicus wrote the pope that he hesitates to publish his book. It reveals that Copernicus and the pope were on speaking terms.

Most of the pupils in our study used sources to validate their interpretations, identify intentions, and justify their answers stemming from a single source. This is demonstrated in particular through the open-ended questions' answers, yet we could not be sure how the pupils were thinking due to the fact that their written answers were quite brief. Concisely put, most of the pupils succeeded in acquiring historical information from at least one source, as well as showing the understanding about the interpretive nature of historical knowledge. However, we noticed that the variation of mastering historical literacy among the participants was huge. Some pupils' answers were far from the standard that is outlined in the assessment criteria at the end of Grade 6. Some others answers would even clearly fulfil the assessment criteria at the end of Grade 8. The first group mentioned above never given thought to the existence of the intentions of the sources' producers. The other group, however, were able to consider the intentions of several producers.

The limitations of our study

The participants in our study were taught by a teacher who follows the guidelines of the national core curriculum. In other words, she teaches multiple-source strategies. As classroom-based interventions have indicated elsewhere, students who are taught multiple-source strategies—as in our case—outperform those who are not accustomed to such an approach (Braasch et al., 2013; Britt & Angliskas, 2002). Cross-textual analysis is not extensively used anyway, but teachers will often provide a single text for students to read (Lee, 2013; Litman et al., 2016; Reisman, 2011). This used to apply to the approach in Finland as well. Today, the majority of history teachers in Finland accept the guidelines provided by the national core curriculum, but there are still teachers who predominantly teach substantive historical knowledge to their students (Ouakrim-Soivio & Rantala, 2017; Rantala, 2012). Presumably, those students are not as capable of handling contrasting sources as the students in our study.

For scrutinising the validity of our test, we tried to find out how 12-year-olds were thinking using the think-aloud sessions, which we implemented with two of the students. In the following transcription, we present a short citation from one of the

sessions. From the citation, it can be seen that the participant notices conflict between the sources.

Question number three. The extract of the textbook says that Copernicus attacked the Church. Can you find something that supports or doesn't support this? Okay... hmmm... I think that the source number three is perhaps against it because Copernicus wrote him... the Pope I mean, like in a mild manner... differently than is said in this source number five... that Copernicus **attacked** [the speaker stresses the word] the Church. So, in this he says it in a gentle way... this is, however, rather against it...²

Using the think-aloud method, we confirmed the way participants handled the documents, as well as the way in which 12-year-olds read historical documents and formulate their own interpretations.

We also verified the validity of the multiple-choice questions with the help of the think-aloud method. It is worth noticing that the participants understood the difference between multiple-choice questions, which usually have one correct answer, and weighted multiple-choice questions consisting of one wrong and several possible answers (with one of them being more compelling than the others). Thus, the task measured pupils' understanding of nuance. As the following excerpt from the think-aloud session reveals, pupils could ponder their answers and did not necessarily immediately choose one of the possible answers.

Question number five. How can you explain that the extract from a textbook gives a different kind of view than the extract from the cardinal's letter on the reaction of the Church against the ideas of Copernicus? Hmmm... It is hardly this option D³, because... in my opinion, this is at least quite important source, otherwise it wouldn't be here. In that case, it isn't the option C. Well... the option A would be quite reasonable, I mean that the view of the textbook is based on many sources which tell that the Catholic Church condemned Copernicus' ideas. Yep, they [the textbook authors] had many sources and majority of them were condemnatory... I mean about these sources produced by the Catholic Church... and then [the textbook authors] thought that the ideas of Copernicus were condemned... and because this source number five is quite tightly written, it couldn't bring out all the views. Not everything showed up in the source number five.

The participants in our study understood the nature of weighted multiple-choice questions. However, this might not generally be the case, since students are used to answering multiple-choice questions that contain just one correct answer. This can lead

² The transcription was originally in Finnish and translated into English for this article.

³ The options were: a) The textbook authors have used different sources that tell that the Church condemned the ideas of Copernicus, and they left out the sources that told the opposite story; b) The cardinal was lying; c) The textbook authors didn't know that some churchmen supported the ideas of Copernicus; d) The textbook authors didn't find the cardinal's letter an important source.

them to choose the first possible answer, thus preventing them from taking the extra time to find the most compelling answer. Using weighted multiple-choice questions is an approach that teachers may need to further familiarise themselves.

Discussion

In Finland, the national core curriculum for basic education defines the learning objectives corresponding to ‘good’ competence (numerical grade 8) in history at the end of Grades 6 and 8. As several studies have been revealed, teachers tend to adjust their grading practices to the general competence level of their class or school, rather than following the national curriculum guideline for student assessment. According to Ouakrim-Soivio (2013), teachers’ evaluation skills vary too much, and teachers should be provided with more coherent assessment criteria. Teachers also experience the lack of assessment guidance from the core curriculum (Ouakrim-Soivio & Rantala, 2017).

Finland is not alone in this problem: in Sweden, Samuelsson and Wendell (2016) found similar difficulties when adapting assessment culture to a new disciplinary history curriculum. When studying standards-based curriculum in history and the realisation of the disciplinary aspects contained therein, they encountered problems in assessing disciplinary knowledge—teachers reinterpreted the curriculum and did not assess the mastery of disciplinary knowledge among students expressly. Similarly, Samuelsson and Wendell discovered that although the pupils involved in their study showed the aspects of disciplinary thinking, the results varied considerably. Lee and Ashby (2000) concluded similar results earlier. These are the same findings we had, and these findings underline the need to develop better assessment instruments and materials for teachers.

The document-based questions have been criticised because they are time-consuming for teachers to correct and wearisome for students to complete (Breakstone, 2014, p. 456). It is true that many page-long tasks can look difficult, both to teachers and students. However, the multiple-choice questions and explicit assessment criteria facilitates teachers’ work. For the students, it mostly involves the adjustment to new assessment methods.

As Counsell (2011) indicates, the changing of a curriculum tradition can take as long as two decades. In Finland, the disciplinary core curriculum for basic school was launched in 2004, and the attitudes of history teachers have already changed to support this new approach. However, the biggest problem, according to the survey made for the history teachers in 2017, seems to be the shortage of sufficient guidance for evaluation. The 537 teachers who answered that survey (about one-fourth of all history teachers at the basic schools and general upper secondary schools in Finland) experienced the lack of assessment guidance from the curriculum (Ouakrim-Soivio & Rantala, 2017). We fear that teachers will once again start to prioritise content mastery over disciplinary literacy if they find the assessment of students’ disciplinary learning too difficult. Therefore, there is an urgent need to train teachers to use disciplinary assessment criteria (see Table 1), along with identifying what level of performance a student must complete in order to fulfil the criteria described in the core curriculum.

Because the insufficient moderation skills of teachers, there is a need for assessment tools that guide teachers. When implementing disciplinary history education in basic education, the Finnish National Board of Education provided such tools at the turn of the millennium. The Board launched an assessment bank where municipalities could order tests for their schools. With the help of those tests, teachers could calibrate their grades nationally. The assessment bank guided teachers to assess their students based on criteria imposed by the National Board of Education, not by means of assessing students against each other. However, the users had to pay for the test. This finally caused the assessment bank to collapse because municipalities could not afford it. At the moment, we are building a similar assessment bank, but with a principle of open access.

It is important to secure the reliability and validity of these tests. To ensure that teachers grade students' answers consistently in open-ended questions (productive tasks), teachers have grade-based criteria and model answers. Using them increases the reliability of the test. The problem of giving full directions of assessment criteria to teachers is connected with the teachers' limited time resources. In order to make sure that teachers will use the assessment guidelines, a test should consist of a limited number of open-ended questions, along with a combination of multiple-choice tasks. Multiple-choice questions are reliable because they produce consistent results. They are able to assess substantive knowledge; however, they do not work well for assessing procedural knowledge (Breakstone, 2014, p. 456; VanSledright, 2014, p. 8). Therefore, we used weighted multiple-choice questions in the Copernicus test. They reveal more about students' cognitive processes and historical thinking (see VanSledright, 2014; VanSledright, 2015). Their validity depends, however, on students' abilities to understand the nature of the tasks and their mastery of the substantive knowledge connected with the question. If students use strategies for answering that are similar to multiple-choice questions and they lack of sufficient content knowledge, the validity of weighted multiple-choice questions is poor. Therefore, teachers who use weighted multiple-choice questions must verify students' substantive historical knowledge and understanding about the nature of the assessment task beforehand, as we did in our study.

The purpose of our study was not to make statistical generalisations, but rather to reveal the type of tasks we can use to assess students' historical literacy. After gathering our study data in November 2015, we have continued to test our task with over one hundred pupils. As stated earlier, our ultimate goal is offer open-access tasks for teachers in order for them to better assess students' historical literacy. In the near future, we will publish this and other tasks we have developed. We are also working to form an online data bank where teachers can leave their ratings. With the help of this data bank, teachers can learn to calibrate their assessments based on a wider context. It is our hope that through these efforts, teachers' assessments will be more objective and the grading more equivocal than is presently the case.

Our paper is connected to the international discussion about teaching and assessing historical thinking at the elementary level. Our findings strengthen the observations done in the CHATA project regarding the heterogeneity of students in the

same age group (Lee & Ashby, 2000). Our study reveals that some of the 12-year-olds are capable of a relatively high level of historical thinking, while some of their peers simultaneously have great difficulties with the same tasks. This was seen in the source exercise: some pupils' capabilities in comparing different sources together were sufficient, while others demonstrated incompetence in handling conflicting sources. Among some scholars, it is believed that elementary pupils are not yet capable of interpreting difficult sources (cf. Brophy & VanSledright, 1997; VanSledright & Afflerbach, 2005). However, by tailoring the sources for the age group in question, even young children are capable of using historical sources (e.g. Barton, 2008; Lee & Ashby, 2000; VanSledright & Afflerbach, 2005). Our paper gives some guidance as to the tasks that can be used with elementary pupils in teaching historical literacy and assessing the learning outcomes.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Academy of Finland under Grant 294491.

References

- Bain, R.B. (2006). Rounding up unusual suspects. Facing the authority hidden in history classroom. *Teachers College Record* 108, 2080–2114.
- Barton, K. C. (2008). “I just kinda know”. Elementary students’ ideas about historical evidence. In L. S. Levstik & K. C. Barton, *Researching History education. Theory, method, and context* (pp. 209–227). New York: Routledge.
- Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students’ evaluation of source features of multiple documents. *Contemporary Educational Psychology*, 38(3), 180–195.
- Breakstone, J. (2014). Try, Try, Try Again: the process of designing new history assessments. *Theory & Research in Social Education*, 42(4), 453–485.
- Britt, M. A., & Angliskas, C. (2002). Improving students’ ability to identify and use source information. *Cognition and Instruction*, 20, 485–522.
- Brookhart, S. M. (2007). Expanding Views About Formative Classroom Assessment: A Review of the Literature. In J. H. McMillan (Ed.), *Formative Classroom Assessment* (pp. 43–62). New York: Teachers College Press.
- Brophy, J., & VanSledright, B. (1997). *Teaching and learning History in elementary schools*. New York: Teachers College Press.
- Counsell, C. (2011). History Teachers as Curriculum Makers: Professional Problem-solving in Secondary School History Education in England. In B. Schüllerqvist (Ed.), *Patterns of Research in Civics, History, Geography and Religious Education* (pp. 53–88). Karlstad, Sweden: Karlstad University Press.

- Downey, M. T., & Long, K. A. (2016). *Teaching for Historical Literacy. Building Knowledge in the History Classroom*. New York: Routledge.
- Hildén, R., Ouakrim-Soivio, N., & Rautopuro, J. (2016). Kaikille ansionsa mukaan? Perusopetuksen päättöarvioinnin yhdenvertaisuus Suomessa [Fair marks for all? Equal and equitable grading in the end of basic education in Finland]. *Kasvatus*, 47(4), 342–357.
- Husbands, C., Kitson, A., & Pendry, A. (2003). *Understanding History Teaching: Teaching and Learning about the Past in Secondary Schools*. Maidenhead-Philadelphia: Open University Press.
- Jakku-Sihvonen, R. (2001). Arvioinnin kehittyminen osaksi julkishallintoa ja koulutuksen ohjausjärjestelmää [the development of assessment as a part of public administration and the steering system of education]. In R. Jakku-Sihvonen, & S. Heinonen. *Johdatus koulutuksen uudistuvaan arviointikulttuuriin* (pp. 9–20). Helsinki: Opetushallitus.
- Jakku-Sihvonen, R., & Heinonen, S. (2001). Opiskelija-arviointi [Students' assessment]. In R. Jakku-Sihvonen, & S. Heinonen. *Johdatus koulutuksen uudistuvaan arviointikulttuuriin* (pp. 78–94). Helsinki: Opetushallitus.
- Kupari, P., Välijärvi, J., Andersson, L., Arfman, I., Nissinen, K., Puhakka, E., & Vettenranta, J. (2013). *PISA12 ensituloksia* [PISA 2012 initial results]. Helsinki: Opetus- ja kulttuuriministeriö.
- Kuusela, J. (2006). *Temaattisia näkökulmia perusopetuksen tasa-arvoon* [Thematic viewpoints on the equality of basic education]. Helsinki: Opetushallitus.
- Lee, M. (2013). Promoting historical thinking using the explicit reasoning text. *Journal of Social Studies Research*, 37(1), 33–45.
- Lee, P., & Ashby, R. (2000). Progression in historical understanding among students ages 7–14. In P. N. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, teaching and learning history: National and international perspectives* (pp 199–222). New York University Press.
- Litman, C., Marple, S., Greenleaf, C., Charney-Sirott, I., Bolz, M.J., Richardson, L.K., Hall, A.H., George, MA. & Goldman, S.R. (2017). Text-based argumentation with multiple sources: A descriptive study of opportunity to learn in secondary English language arts, history, and science. *Journal of the Learning Sciences*, 26(1), 79–130.
- National Board of Education (1999). *A Framework for Evaluating Educational Outcomes in Finland*. Helsinki: National Board of Education. Retrieved from https://karvi.fi/app/uploads/2014/09/OPH_0499.pdf.
- National Board of Education (2014). *National Core Curriculum for Basic Education 2014*. Helsinki: The Finnish National Board of Education.
- Nokes, J. D. (2013). *Building students' historical literacies: Learning to read and reason with historical texts and evidence*. New York: Routledge.
- Perfetti, C. A., Britt, M. A., Rouet, J.-F., Georgi, M. C. & Mason, R. A. (1994). How students use texts to learn and reason about historical uncertainty. In M. Carretero & J.F. Voss (Eds.), *Cognitive and Instructional Processes in History and Social Sciences* (pp. 257–283). Hillsdale: Lawrence Erlbaum.

- OECD (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*, PISA. Paris: OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/9789264266490-en>.
- Ouakrim-Soivio, N. (2013). *Toimivatko päättöarvioinnin kriteerit? Oppilaiden saamat arvosanat ja Opetushallituksen oppimistulosten seuranta-arvioinnit koulujen välisten osaamiserojen mittareina* [Are the criteria for final grading functioning? Student grades and the reviews on learning outcomes by the National Board of Education as indicators of between-school differences]. Helsinki: Finnish National Board of Education.
- Ouakrim-Soivio, N., & Kuusela, J. (2012). *Historian ja yhteiskuntaopin oppimistulokset perusopetuksen päättövaiheessa 2011*. [The learning outcomes in history and social studies on the final (9th) Grade of compulsory basic education in 2011]; Helsinki: Finnish National Board of Education.
- Ouakrim-Soivio, N. & Rantala, J. (2017). The results of the survey done with the history teachers in Finland in 2017. In the author's possession.
- Perfetti, C. A., Britt, M. A., Rouet, J.-F., Georgi, M. C., & Mason, R. A. (1994). How students use texts to learn and reason about historical uncertainty. In M. Carretero, & J. F. Voss (Eds.), *Cognitive and Instructional Processes in History and Social Sciences* (pp. 257–283). Hillsdale, N.J.: Lawrence Erlbaum.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.
- Rantala, J. (2012). How Finnish Adolescents Understand History: Disciplinary Thinking in History and Its Assessment Among 16-Year-Old Finns. *Education Sciences*, 2(4), 193–207.
- Rantala, J., & van den Berg, M. (2015). Finnish high school and university students' ability to handle multiple source documents in history. *Historical Encounters: A journal of historical consciousness, historical cultures, and history education*, 2(1), 70–88.
- Rantala, J., & Khawaja A. (2016). Historian tekstitaitojen arvioiminen peruskoulun kuudesluokkalaisilla [Assessing historical literacy among elementary pupils]. In H.-M. Pakula, E. Kouki, H. Silfverberg, & E. Yli-Panula (Eds.), *Uudistuva ja uusiutuva ainedidaktiikka* (pp. 77–94). Helsinki: Ainedidaktinen tutkimusseura.
- Reisman, A. (2011). The “document-based lesson”: Bringing disciplinary inquiry into high school history classrooms with adolescent struggling readers. *Journal of Curriculum Studies*, 44(2), 233–264.
- Reisman, A. (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction*, 30(1), 86–112.
- Samuelsson, J., & Wendell, J. (2016). Historical thinking about sources in the context of a standards-based curriculum: a Swedish case. *The Curriculum Journal*, 27(4), 479–499.
- Seixas, P. & Colyer, J. (2011). Continuing the Momentum – Benchmarks of historical thinking. A report on to Toronto meeting February 10–12, 2011. Retrieved from <http://historicalthinking.ca/sites/default/files/files/docs/Continuing%20the%20Momentum.pdf>.

- Seixas, P., Gibson, L. & Ercikan, K. (2015). A Design Process for Assessing Historical Thinking: The Case of a One-Hour Test. In K. Ercikan, & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 102–116). New York: Routledge.
- Seixas, P. & Morton, T. (2013). *The Big Six Historical Thinking Concepts*. Toronto: Nelson.
- Seixas, P. & Ercikan, K. (2015). Introduction: The new shape of history assessment. In K. Ercikan, & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 1–13). New York: Routledge.
- Stahl, A. & Shanahan, C. (2004). Learning to think like a historian. Disciplinary knowledge through critical analysis of multiple documents. In T. L. Jetton & J. A. Dole (Eds.), *Adolescents literacy research and practice* (pp. 94–118). New York: Guilford.
- Thorp, R. & Törnqvist, E. (2017). Young Children's Historical Consciousness: A Swedish Case Study. In *International Journal of Research on History Didactics, History Education and History Culture* (pp. 215–234). Schwalbach: Wochenschau Verlag.
- VanSledright, B. (2002). *In Search of America's Past. Learning to read history in elementary school*. New York and London: Teachers College Press.
- VanSledright, B. (2014). *Assessing Historical Thinking & Understanding. Innovative Designs for New Standards*. New York: Routledge.
- VanSledright, B. (2015). Assessing for Learning in the History Classroom. In K. Ercikan & P. Seixas (Eds.), *New Directions in Assessing Historical Thinking* (pp. 75–88). New York: Routledge.
- VanSledright, B., & Afflerbach, P. (2005). Assessing the status of historical sources: An exploratory study of eight US elementary students reading documents. In R. Ashby, P. Gordon, & P. Lee (Eds.), *Understanding history: Recent research in history education* (pp. 1–20). London: RoutledgeFalmer.
- VanSledright, B., & Frankes, L. (2000). Concept- and Strategic-Knowledge Development in Historical Study: A Comparative Exploration in Two Fourth-Grade Classrooms. *Cognition and Instruction*, 18(2), 239–283.
- Vettenranta, J., Välijärvi, J., Ahonen, A., Hautamäki, J., Hiltunen, J., Leino, K., Lähteenmäki, S., Nissinen, K., Nissinen, V., Puhakka, E., Rautopuro, J., & Vainikainen, M.-P. (2016). *PISA15 ensituloksia* [PISA 2015 initia results]. Helsinki: Opetus- ja kulttuuriministeriö.
- Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology* 83(1), 73–97.
- Wineburg, S. (2001). *Historical thinking and other unnatural acts*. Philadelphia: Temple University Press.
- Wineburg, S. & Reisman, A. (2015). Disciplinary literacy in history. A toolkit for digital citizenship. *Journal of Adolescent & Adult Literacy* 58(8), 636–639.